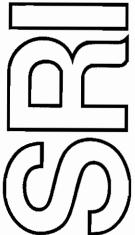# TWO PRINCIPLES OF PARSE PREFERENCE

Technical Note 483

April 18, 1990

By: Jerry R. Hobbs, Sr. Computer Scientist
        and
    John Bear, Computer Scientist
    Artificial Intelligence Center
    Computing and Engineering Sciences Division

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **18 APR 1990**	2. REPORT TYPE	3. DATES COVERED **00-04-1990 to 00-04-1990**
4. TITLE AND SUBTITLE **Two Principles of Parse Preference**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT   **Approved for public release; distribution unlimited**
13. SUPPLEMENTARY NOTES
14. ABSTRACT
15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**11**	

Two Principles of Parse Preference

Jerry R. Hobbs and John Bear
Artificial Intelligence Center
SRI International

1   Introduction

The DIALOGIC system for syntactic analysis and semantic translation has been under development for over ten years, and during that time it has been used in a number of domains in both database interface and message-processing applications. In addition, it has been tested on a number of sentences of linguistic interest. Built into the system are facilities for ranking parses according to syntactic and selectional considerations, and over the years, as various kinds of ambiguity have become apparent, heuristics have been devised for choosing the preferred parses. Our aim in this paper is first to present a compendium of many of these heuristics and secondly to propose two principles that seem to underlie the heuristics. The first will be useful to researchers engaged in building grammars of similarly broad coverage. The second is of psychological interest and may be a guide for estimating parse preferences for newly discovered ambiguities for which we lack the experience to decide among on a more empirical basis.

The mechanism for implementing parse preference heuristics is quite simple. Terminal nodes of a parse tree acquire a score (usually 0) from the lexical entry for the word sense. When a nonterminal node of a parse tree is constructed, it is given an initial score which is the sum of the scores of its child nodes. Various conditions are checked during the construction of the node and, as a result, a score of 20, 10, 3, -3, -10, or -20 may be added to the initial score. The score of the parse is the score of its root node. The parses of ambiguous sentences are ranked according to their scores. Although simple, this method has been very successful. In this paper, however, rather than describe the heuristics in terms this detailed, we will describe them in terms of the preferences among the alternate structures that motivated our scoring schemes.

While these heuristics have arisen primarily through our everyday experience with the system, we have done small empirical studies by hand on some of the ambiguities, using several different kinds of text, including some from the Brown corpus and some transcripts of spoken dialogue. We have counted the number of occurrences of potentially ambiguous constructions that were in accord with our claims, and the number of occurrences that were not. Some of the constructions were impossible to find, not only because they occur so rarely but also because many are very difficult for anyone except a dumb parser to spot. But in every case where we found examples, the numbers supported our claims. We present our preliminary findings below for those cases where we have begun to accumulate a nontrivial number of examples.

2   Brief Review of the Literature

Most previous work on parse preferences has concerned itself with the most notorious of the ambiguities—the attachment ambiguities of postmodifiers. Among the first linguists to address this problem was Kimball (1973). He proposed several processing principles in an attempt to account for why certain readings of ambiguous sentences were more salient than others. Two of these principles were Right Association and Closure.

In the late 1970s and early 1980s there was a great deal of work among linguists and psycholinguists (e.g. Frazier and Fodor, 1979; Wanner and Maratsos, 1978; Marcus, 1979; Church, 1980; Ford, Bresnan, and Kaplan, 1982) attempting to refine Kimball's initial analysis of syntactic bias and proposing their own principles govering attachment. Frazier and Fodor proposed the principles of Minimal Attachment and Local Association. Church proposed the A-over-A Early Closure Principle; and Ford, Bresnan and Kaplan introduced the notions of Lexical Preference and Final Arguments.

The two ideas that dominated their hypotheses and discussions were Right Association, which says roughly that postmodifiers prefer to be attached to the nearest previous possible head, and a stronger principle stipulating that argument interpretations are favored over adjunct interpretations. This latter principle is implied by Frazier and Fodor's Minimal Attachment and also by Ford, Bresnan and Kaplan's Lexical Preference.

In recent computational linguistics, Shieber and Pereira (Shieber, 1983; Pereira, 1985) proposed a shift-reduce parser for parsing English, and showed that Right Association was equivalent to preferring shifts over reductions, and that Minimal Attachment was equivalent to favoring the longest possible reduction at each point.

More recently, there have been debates, for example, between Schubert (1984, 1986) and Wilks et al. (1985), about the interaction of syntax with semantics and the role of semantics in disambiguating the classical ambiguities.

We take it for granted that, psychologically, syntax, semantics, and pragmatics interact very tightly to achieve disambiguation. In fact, in other work (Hobbs et al., 1988), we have proposed an integrated framework for natural language processing that provides for this tight interaction. However, in this paper, we are considering only syntactic factors. In the semantically and pragmatically unsophisticated systems of today, these are the most easily accessible factors, and even in more sophisticated systems, there will be examples that semantic and pragmatic factors alone will fail to disambiguate.

The two principles we propose may be viewed as generalizations of Minimal Attachment and Right Association.

3   Most Restrictive Context

The first principle might be called the Most Restrictive Context principle. It can be stated as follows:

> Where a constituent can be placed in two different structures, favor the structure that places greater constraints on allowable constituents.

For example, in

John looked for Mary.

"for Mary" can be interpreted as an adverbial signaling the beneficiary of the action or as a complement of the verb "look". Since virtually any verb phrase can take an adverbial whereas only a very few verbs can take a "for" prepositional phrase as its complement, the latter interpretation has the most restrictive context and therefore is favored.

A large number of preferences among ambiguities can be subsumed under this principle. They are enumerated below.

1. As in the above example, favor argument over adverbial intepretations for post-modifying prepositional phrases where possible. Thus, whereas in

John cooked for Mary.

"for Mary" is necessarily an adverbial, in "John looked for Mary" it is taken as a complement. Subsumable under this heuristic is the preference of "by" phrases after passives to indicate the agent rather than a location. This heuristic, together with the next type, constitutes the traditional Minimal Attachment principle. This heuristic is very strong; of 47 occurrences examined, all were in accord with the heuristic.

2. Favor arguments over mere modifiers. Thus, in

John bought a book from Mary.

the favored interpretation is "bought from Mary" rather than "book from Mary". Where the head noun is also subcategorized for the preposition, as in,

John sold a ticket to the theater.

this principle fails to decide among the readings, and the second principle, described in the next section, becomes decisive.

This principle was surprisingly strong, but perhaps for illegitimate reasons. Of 75 potential ambiguities, all but one were in accord with the heuristic. The one exception was

HDTV provides television images with finer detail than current systems.

and even this is a close call. However, it is often very uncertain whether we should say verbs, nouns, and adjectives subcategorize for a certain preposition. For example, does "discussion" subcategorize for "with" and "about"? We are likely to say so when it yields the right parse and not to notice the possibility when it would yield the wrong parse. So our results here may not be completely unbiased.

3. Favor complement interpretations of infinitives over purpose adverbial interpretations. In

John wants his driver to go to Los Angeles.

the preferred interpretation has only the driver and not John going to Los Angeles.

Of 44 examples of potential ambiguities of this sort that we found, 41 were complements and only 3 were purpose adverbials. Even these three could have been eliminated with the simplest selectional restrictions. One example was the following

3

He pushed aside other business to devote all his time to this issue.

which could have been parsed analogously to

He pushed strongly all the young researchers to publish papers on their work.

A particularly intriguing example, remembering that "provide" can be ditransitive, is the following:

That is weaker than what the Bush administration needs to provide the necessary tax revenues.

4. Favor the attachment of temporal prepositional phrases to verbs or event nouns. In the preferred reading of

John saw the President during the campaign.

the seeing was during the campaign, since "President" is not an event noun. In the preferred reading of

The historian described the demonstrations during Gorbachev's visit.

the demonstrations are during the visit. This case can be considered an example of Minimal Attachment if we assume that all verbs and event nouns have potential temporal arguments. Of 74 examples examined, 66 were in accord with this heuristic. Two that did not involved the phrase "business since August 1".

5. Favor adverbial over object interpretations of temporal and measure noun phrases. Thus, in

John won one day in Hawaii.

"one day in Hawaii" is preferentially the time John won and not his prize. In

John walked 10 miles.

"10 miles" is a measure of how far he walked, not what he walked. This is an example of Most Restrictive Context because noun phrases, based on syntactic criteria alone, can always be the object of a transitive verb, whereas only temporal and measure noun phrases can function as adverbials. This case is interesting because it runs counter to Minimal Attachment. Here arguments are *disfavored.*

Of fifteen examples we found of such ambiguities, eleven agreed with the heuristic. The reason for the large percentage of examples that did not is that sports articles were among those examined, and they contained sentences like

Smith gained 1240 yards last season.

This illustrates the hidden dangers in genre selection.

6. Favor temporal nouns as adverbials over compound nominal heads. The latter interpretation is possible, as seen in

Is this a CSLI Thursday?

But the preferred reading is the temporal one that is most natural in

I saw the man Thursday.

7. Favor "that" as a complementizer rather than as a determiner. Thus, in

I know that sugar is expensive.

we are probably not referring to "that sugar". This is a case of Most Restrictive Context because the determiner "that" can appear in any noun phrase, whereas the complementizer "that" can occur only after a small number of verbs. This is a heuristic we suspect everyone who has built a moderately large grammar has implemented, because of the frequency of the ambiguity.

8. An initial "there" is interpreted as an existential, where possible, rather than as a locative. We interpret

There is a man in the room.

as an existential declarative sentence, rather than as an utterance with an initial locative. Locatives can occur virtually anyplace, whereas the existential "there" can occur in only a very small range of contexts. Of 30 occurrences examined, 29 were in accord with the heuristic. The one exception was

There, in the midst of all those casinos, is Trump's Taj Mahal.

9. Favor predeterminers over separate noun phrases. In

Send all the money.

the reading that treats "all the" as a complex determiner is favored over the one that treats "all" as a separate complete noun phrase in indirect object position. There are very many fewer loci for predeterminers than for noun phrases, and hence this is also an example of Most Restrictive Context.

10. Favor prepositional lexical adverbs over separate adverbials. Thus, in

John did the job precisely on time.

we favor "precisely" modifying "on time" rather than "did the job". Very many fewer adverbs can function as prepositional modifiers than can function as verbal or sentential adverbs. Of 28 occurrences examined, all but one were in accord with the heuristic. The one was

Who is going to type this all for you?

11. Group numbers with prenominal unit nouns but not with other prenominal nouns. For example, "10 mile runs" are taken to be an indeterminate number of runs of 10 miles each rather than as exactly 10 runs of a mile each. Other nouns can function the same way as unit nouns, as in "2 car garages", but it is vastly more common to have the number

5

attached to the head noun instead, as in "5 wine glasses". Virtually any noun can appear as a prenominal noun, whereas only unit nouns can appear in the adjectival "10-mile" construction. Hence, for unit nouns this is the most restrictive context. While other nouns can sometimes occur in this context, it is only through a reinterpretation as a unit noun, as in "2 car garages".

12. Disfavor headless structures. Headless structures impose no constraints, and are therefore never the most restrictive context, and thus are the least favored in cases of ambiguity. An example of this case is the sentence

John knows the best man wins.

which we interpret as a concise form of

John knows (that) the best man wins.

rather than as a concise form of

John knows the best (thing that) man wins ().

4 Attach Low and Parallel

The second principle might be called the Attach Low and Parallel principle. It may be stated as follows:

Attach constituents as low as possible, and in parallel with other constituents if possible.

The cases subsumed by this principle are quite heterogeneous.

1. Where not overridden by the Most Restrictive Context principle, favor attaching postmodifiers to the closest possible site, skipping over proper nouns. Thus, where neither the verb nor the noun is subcategorized for the preposition, as in

John phoned a man in Chicago.

or where *both* the verb and the noun are subcategorized for the preposition, as in

John was given a book by a famous professor.

the noun is favored as the attachment point, since that is the lowest possible attachment point in the parse tree. This case is just the traditional Right Association.

The subcase of prepositional phrases with "of" is significant enough to be mentioned separately. We might say that every noun is subcategorized for "of" and that therefore "of" prepositional phrases are nearly always attached to the immediately preceding word. Of 250 occurrences examined, 248 satisfied this heuristic, and of the other two

Since the first reports broke of the CIA's activities, ...

He ordered the destruction two years ago of some records.

the second would not admit an incorrect attachment in any case.

We examined 148 instances of this case not involving "of", temporal prepositional phrases, or prepositions that are subcategorized for by possible attachment points. Of these, 116 were in accord with the heuristic and 32 were not. An example where this heuristic failed was

They abandoned hunting for food production.

For a significant number of examples (34), it did not matter where the attachment was made. For instance, in

John made coffee for Mary.

both the coffee and the making are for Mary. We counted these cases as being in accord with the heuristic, since the heuristic would yield a correct interpretation.

This is perhaps the place to present results on two very simple algorithms. The first is to attach prepositional phrases to the closest possible attachment point, regardless of other considerations. Of 251 occurrences examined, 125 attached to the nearest possibility, 109 to the second nearest, 14 to the third, and 3 to the fourth, fifth, or sixth. This algorithm is not especially recommended.

The second algorithm is to attach to the nearest possible attachment point that subcategorizes for the preposition, if there is such, assuming verbs and event nouns to subcategorize for temporal prepositional phrases, and otherwise to attach to the nearest possible attachment point. This is essentially a summary of our heuristics for prepositional phrases. Of 297 occurrences examined, this yielded the right answer on 256 and the wrong one on 41.

2. Favor preprepositional readings of measure phrases over readings as separate adverbials. Thus, in

John walked 10 miles into the forest.

we preferentially take "10 miles" as modifying "into the forest" rather than "walked", so that John is now 10 miles from the edge of the forest, rather than merely somewhere in the forest but 10 miles from his starting point. Since the preposition occurs lower in the parse tree than the verb, this is an example of Attach Low and Parallel. Note that this is a kind of "Left Association".

3. Coordinate "both" with "and", if possible, rather than treating it as a separate determiner. In

John likes both intelligent and attractive women.

the interpretation in which there are exactly two women who are intelligent and attractive is disfavored. Associating "both" with the coordinated adjectives rather than attaching it to the head noun is attaching it lower in the parse tree.

4. Distribute prenominal nouns over conjoined head nouns. In "oil sample and filter", we mean "oil sample and oil filter". A principle of Attach Low would not seem to be decisive in this case. Would it mean that we attach "oil" low by attaching it to "sample"

or that we attach "and filter" low by attaching it to "sample". It is because of examples like this (and the next case) that we propose the principle Attach Low *and Parallel*. We favor the reading that captures the parallelism of the two head nouns.

5. Distribute determiners and noun complements over conjoined head nouns. In "the salt and pepper on the table", we treat "salt" and "pepper" as conjoined, rather than "the salt" and "pepper on the table". As in the previous case, where we have a choice of what to attach low, we favor attaching parallel elements low.

6. Favor attaching adjectives to head nouns rather than prenominal nouns. We take "red boat house" to refer to a boat house that is red, rather than to a house for red boats. Like all of our principles, this preference can be overridden by semantics or convention, as in "high stress job". Here again we could interpret Attach Low as telling us to attach "red" to "boat" or to attach "boat" to "house". Attach Low and Parallel tells us to favor the latter.

5   Interaction and Overriding

There will of course be many examples where both of our principles apply. In the cases that occur with some frequency, in particular, the prepositional phrase attachment ambiguities, it seems that the Most Restrictive Context principle dominates Attach Low and Parallel. It is unclear what the interactions between these two principles should be, more generally.

These principles can be overridden by more than just semantics and pragmatics. Commas in written discourse and pauses in spoken discourse (see Bear and Price, 1990, on the latter) often function to override Attach Low and Parallel, as in

John phoned the man, in Chicago.

Specify the length, in bits, of a word.

It is the phoning that is in Chicago, and the specification is in bits while the length is of a word. Similarly, commas and pauses can override the Most Restrictive Context principle, as in

John wants his driver, to go to Los Angeles.

Here we prefer the purpose adverbial reading in which John and the driver both are going to Los Angeles.

6   Cognitive Significance

The analysis of parse preferences in terms of these two very general principles is quite appealing, and more than simply because they subsume a great many cases. They seem to relate somehow to deep principles of cognitive economy. The Most Restrictive Context principle is a matter of taking all of the available information into account in constructing interpretations. The "Low" of Attach Low and Parallel is an instance of a general cognitive heuristic to interpret features of the environment as locally as possible. The "Parallel" exemplifies a general cognitive heuristic to see similarity wherever possible, a heuristic that promotes useful generalizations.

8

Acknowledgements

References

[1] Bear, John, and Jerry Hobbs, 1988. "Localizing Expression of Ambiguity", *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas, pp. 235-241.

[2] Bear, John, and Patti Price, 1990. "Prosody, Syntax and Parsing", *Proceedings*, 28th Annual Meeting of the Association for Computational Linguistics, Pittsburgh, Pennsylvania.

[3] Church, Kenneth, 1980. "On Memory Limitations in Natural Language Processing", MIT Technical Report MIT/LCS/TR-245.

[4] Ford, Marylyn, Joan Bresnan, and Ronald Kaplan, 1982. "A Competence-Based Theory of Syntactic Closure," in J. Bresnan (Ed.) *The Mental Representation of Grammatical Relations*, MIT Press: Cambridge, Massachusetts.

[5] Frazier, Lyn and Janet Fodor, 1979. "The Sausage Machine: A New Two-Stage Parsing Model", *Cognition*, Vol. 6, pp. 291-325.

[6] Hobbs, Jerry R., Mark Stickel, Paul Martin, and Douglas Edwards, 1988. "Interpretation as Abduction", *Proceedings, 26th Annual Meeting of the Association for Computational Linguistics*, pp. 95-103, Buffalo, New York, June 1988.

[7] Kimball, John, 1973. "Seven Principles of Surface Structure Parsing in Natural Language", Cognition Vol. 2, No. 1, pp. 15-47.

[8] Marcus, Mitchel, 1980. *A Theory of Syntactic Recognition for Natural Language*, MIT Press: Cambridge, Massachusetts.

[9] Pereira, Fernando, 1985. "A New Characterization of Attachment Preferences," in D. Dowty et al. (Eds.) *Natural Language Processing*, Cambridge University Press: Cambridge, England.

[10] Schubert, Lenhart, 1984. "On Parsing Preferences", *Proceedings, COLING 1984*, Stanford, California, pp. 247-250.

[11] Schubert, Lenhart, 1986. "Are There Preference Trade-offs in Attachment Decisions?" *Proceedings, AAAI 1986*, Philadelphia, Pennsylvania, pp. 601-605.

[12] Shieber, Stuart, 1983. "Sentence Disambiguation by a Shift-Reduce Parsing Technique", *Proceedings, IJCAI 1983*, Washington, D.C., pp. 699-703.

[13] Wanner, Eric, and Michael Maratsos, 1978. "An ATN Approach to Comprehension," in Halle, Bresnan, and Miller (Eds.) *Linguistic Theory and Psychological Reality.* MIT Press: Cambridge, Massachusetts.

[14] Wilks, Yorick, Xiuming Huang, and Dan Fass, 1985. "Syntax, Preference and Right Attachment", *Proceedings, IJCAI 1985*, Los Angeles, California, pp. 779-784.